

# Selection of functional 2A sequences within foot-and-mouth disease virus; requirements for the NPGP motif with a distinct codon bias

**Kjær, Jonas; Belsham, Graham J.**

*Published in:*
R N A

*Link to article, DOI:*
10.1261/rna.063339.117

*Publication date:*
2018

*Document Version*
Peer reviewed version

# Selection of functional 2A sequences within foot-and-mouth disease virus; requirements for the NPGP motif with a distinct codon bias

Jonas Kjær[1], Graham J. Belsham[1].

DTU National Veterinary Institute, Technical University of Denmark, Lindholm, DK-4771 Kalvehave, Denmark

Short title: Codon bias for NPGP motif in FMDV 2A

Keywords: Picornavirus; Synonymous codon; Codon bias; translation

Corresponding author: Graham J. Belsham, grbe@vet.dtu.dk

## Abstract

Foot-and-mouth disease virus (FMDV) has a positive-sense ssRNA genome including a single, large, open reading frame. Splitting of the encoded polyprotein at the 2A/2B junction is mediated by the 2A peptide (18 residues long) which induces a non-proteolytic, co-translational, "cleavage" at its own C-terminus. A conserved feature among variants of 2A is the C-terminal motif $N^{16}P^{17}G^{18}/P^{19}$ where $P^{19}$ is the first residue of 2B. It has been shown previously that certain amino acid substitutions can be tolerated at residues $E^{14}$, $S^{15}$ and $N^{16}$ within the 2A sequence of infectious FMDVs but no variants at residues $P^{17}$, $G^{18}$ or $P^{19}$ have been identified. In this study, using highly degenerate primers, we analysed if any other residues can be present at each position of the NPG/P motif within infectious FMDV. No alternative forms of this motif were found to be encoded by rescued FMDVs after 2, 3 or 4 passages. However, surprisingly, a clear codon preference for the wt nucleotide sequence encoding the NPGP motif within these viruses was observed. Indeed, the codons selected to code for $P^{17}$ and $P^{19}$ within this motif were distinct; thus the synonymous codons are not equivalent.

## Introduction

Foot-and-mouth disease virus (FMDV) is the prototypic member of the genus *Aphthovirus* within the family *Picornaviridae*. This virus is the causative agent of the highly contagious and economically important disease of cloven-hoofed animals, foot-and-mouth disease. The positive-sense ssRNA genome of around 8400 nt includes a single, large, open reading frame (ORF), ca. 7000 nt, encoding a polyprotein (Belsham 2005). The full-length viral polyprotein is never observed since it is rapidly processed during and after synthesis mainly by the virus-encoded proteases (primarily $3C^{pro}$) to produce 15 distinct mature proteins plus multiple precursors (reviewed in Martinez-Salas and Belsham 2017). Interestingly, FMDV, like many (but by no means all) other picornavirus (e.g. cardioviruses, erboviruses, teschoviruses etc.) employs a co-translational, protease-independent mechanism for the "cleavage" of the polyprotein at the 2A/2B junction (the boundary between the capsid proteins and the non-structural proteins) (Donnelly et al. 2001a). This mechanism has been referred-to as "ribosomal skipping" or, alternatively, "stop-carry on" or "StopGo" (Atkins et al. 2007; Doronina et al. 2008; Donnelly et al. 2001a; Tulloch et al. 2017). The 2A peptide lacks characteristic protease motifs and only mediates "cleavage" during translation. It has been demonstrated that the 2A sequence is able to mediate "cleavage" in all eukaryotic translation systems tested whereas a number of artificial polyproteins containing this sequence have been examined in prokaryotic systems and no detectable cleavage products were observed (Donnelly et al., 1997).

The 2A peptide contains a highly conserved $D^{12}(V/I)E(S/T)NPG_{2A}{}^{\downarrow}P^{19}{}_{2B}$ motif at its C-terminus which is critical for its function (Ryan and Drew 1994; Donnelly et al. 1997). This motif, together with upstream amino acids, is believed to interact with the ribosomal exit tunnel. This prevents the formation of a peptide bond between the C-terminal glycine ($G^{18}$) of 2A and the N-terminal proline of 2B, referred to here, as $P^{19}$ since it is an important part of

the cleavage mechanism (see also Donnelly et al. 2001a; Ryan et al. 1999). However, remarkably, protein synthesis continues without the requirement for a re-initiation event.

Investigations into the activity of the 2A sequence have mainly been performed using *in vitro* experiments. Typically, these have either used mRNAs with single ORFs encoding artificial polyproteins comprising two reporter proteins linked via the 2A peptide (Ryan et al. 1991; Ryan and Drew 1994; Donnelly et al. 2001b) or by expressing cDNAs encoding a truncated viral polyprotein including the StopGo coding region (Palmenberg et al. 1992). Alterations to the conserved $D^{12}$(V/I)E(S/T)NP$G_{2A}{}^{\downarrow}P^{19}{}_{2B}$ motif reduced or abrogated the StopGo function (Donnelly et al. 2001b; Sharma et al. 2012), thereby showing that these amino acids are important for the correct StopGo "cleavage". Furthermore, Hahn and Palmenberg (1996) demonstrated that alterations to this motif also influenced the viability of encephalomyocarditis virus (EMCV, a cardiovirus) as they resulted in lethal phenotypes. Subsequently, Loughran et al. (2013) reported a similar observation for FMDV, as modification of the $S^{15}$NP$G^{18}{}_{2A}{}^{\downarrow}P^{19}{}_{2B}$ sequence to $S^{15}$NP$L^{18}{}_{2A}{}^{\downarrow}V^{19}{}_{2B}$ or $S^{15}$NP$A^{18}{}_{2A}{}^{\downarrow}P^{19}{}_{2B}$ also gave rise to a lethal phenotype.

However, recently, certain amino acid substitutions (e.g. 2A $S^{15}$ to F/I and 2A $N^{16}$ to H) that have been shown to severely (60-70%) impair "cleavage" at the 2A/2B junction, using *in vitro* assays (Donnelly et al. 2001b), have been found to be tolerated within infectious FMDVs (Kjær J and Belsham GJ, submitted). In contrast, other substitutions (e.g. $P^{19}$ to A and $P^{19}$ to G) that inhibit cleavage more severely (by 89-100%) *in vitro*, were not found within rescued viruses. Indeed, viruses rescued from these mutant transcripts had sequences that exactly matched the wt sequence (i.e. the rescued viruses were not mutant). In these studies, we also determined a critical role for the StopGo mechanism for the overall level of replication/translation of FMDV RNA. FMDV replicons with a defective 2A sequence had a

markedly lower replication efficiency compared to the wt replicon (Kjær J and Belsham GJ, submitted).

It is, therefore, apparent that some amino acid substitutions can be tolerated within the FMDV 2A peptide whereas other changes are not compatible with viability. To identify if any alternative residues can be accepted within the critical $N^{16}P^{17}G^{18}{}_{2A}{}^{\downarrow}P^{19}{}_{2B}$ motif, degenerate sequences, encoding all possible amino acid substitutions at each of these positions individually, were introduced into a full-length FMDV cDNA, as used previously (Gullberg et al. 2013; Kristensen et al. 2017). In principle, this should result in the production of RNA transcripts encoding 2A peptides with a wide spectrum of "cleavage" activities. This was achieved by generating a large pool of plasmids, using site-directed mutagenesis with highly degenerate oligonucleotides, to change each of the individual codons corresponding to the amino acid residues within this conserved motif to NNN (where N is a mixture of all 4 bases). Using each pool of plasmids, RNA transcripts were prepared, *in vitro*, and introduced into baby hamster kidney (BHK) cells. Infectious viruses were rescued and characterized.

## Results and Discussion

The expected generation of a pool of StopGo cDNA mutants that could potentially result in all possible single amino substitutions in place of the $N^{16}$, $P^{17}$, $G^{18}$ and $P^{19}$ residues (see Figure 1A) was analysed by sequencing (see Figure 1B). The heterogeneity at the expected positions was clear in each case (this does not prove that each of the possible codons was present but indicates it is likely).

Full-length RNA transcripts were produced, *in vitro*, and introduced into BHK cells. Infectious virus was generated and passaged in fresh cells. RNA was then extracted from the virus harvests and the sequence encoding the 2A peptide was amplified by RT-PCR. The

pool of amplicons was introduced into the pCR-XL-TOPO vector and then the sequence of the inserts in 20 individual colonies was determined for each virus harvest. It was found that all of the rescued viruses analysed after passages p2, p3 and p4 encoded the wt amino acid sequence at the NPGP motif in 2A. Interestingly, the complete spectrum of the possible synonymous codons for each of the residues $N^{16}$, $P^{17}$, $G^{18}$ and $P^{19}$ was present in the rescued viral genomes at p2 (see Table 1). These results indicated that the approach had indeed generated a diverse pool of codons within the viruses. Furthermore, the very restricted range of nucleotide sequences encoding 2A observed within the rescued viruses strongly suggests that the specific amino acid sequence (NPGP), encoded by these nucleotide sequences, is critical for FMDV viability.

However, it was also apparent that the utilization of the different codons for the conserved amino acid residues varied. At p2, 55% of the sequences analysed had the wt codon for residue $N^{16}$ (AAC) while the synonymous AAT codon was present in the remaining 45% of the rescued sequences. In the subsequent passages, the proportion of the AAC codon within the sequences increased to 75% and 95% by p3 and p4 respectively while the incidence of the AAT codon declined (Table 1). For residue $P^{17}$, at p2, the codon CCT was present in 55% of the colonies analysed and increased to 100% by p4. Each of the three other possible codons for $P^{17}$ (CCC, CCA and CCG) were also observed at p2 but each declined as the wt codon became dominant. For residues $G^{18}$ and $P^{19}$, the wt codons (GGG and CCC respectively) were in the minority (10 or 20%) at p2 and each of the synonymous codons were also present. However, interestingly, by p3 the wt codons had markedly increased to 50% abundance and by p4 were dominant (≥90% abundance). For $G^{18}$, the GGA codon was the most abundant at p2 but declined during further passages to be only 10% of the sequences at p4. Similarly, for $P^{19}$ the CCT codon was present in 50% of the sequences at p2 but declined to just 5% by p4. Strikingly, by p3, the wt codon was present in 50-75% of the population at each of the 4

residues and by p4 the wt codon was present in 90-100% of the virus population in each case (Table 1). Thus, it appears that selection occurs for the wt nucleotide sequence during passage of the rescued viruses in cell culture.

The wt GGGCCC nt sequence encoding residues $G^{18}$ and $P^{19}$ is recognized in DNA by the restriction enzyme *ApaI* (see Figure 1A). Hence, it was possible to deplete the cDNA amplicons generated by RT-PCR, of the wt sequence from the rescued viruses by digesting them with *ApaI* prior to the cloning step (it was anticipated that this should enhance the detection of non-wt nucleotide sequences). The residual, full-length, 650bp amplicons were inserted into the pCR-XL-TOPO vector, as described above, and the plasmid DNA from individual colonies was sequenced. As expected, the wt codons for $G^{18}$ and $P^{19}$ were no longer observed in the cloned fragments (Table 2) and the $G^{18}$ (GGA) and $P^{19}$ (CCT) codons were predominant in these enriched populations. These results are consistent with those obtained without the *ApaI* digestion (since the GGA and CCT codons were also present in 50% of the fragments at p2 without this treatment, see Table 1) but clearly the apparent abundance of these non-wt codons is enhanced following the *ApaI* digestion (Table 2), as anticipated. The enrichment for non-wt sequences did not result in the detection of codons for alternative amino acids within the virus population. It had been anticipated that some amino acid substitutions at residue $N^{16}$ might be rescued since a mutant (with $N^{16}$ changed to H) has been shown to be viable (Kjær J and Belsham GJ, submitted) but, presumably, it was outcompeted by the wt virus.

It is interesting to note that the $G^{18}$ (GGA) and $P^{19}$ (CCT) codons have previously been found to be the second most abundant codons found in FMDV genomes from all seven serotypes (see Gao et al. 2014). This comparison of FMDV sequences also indicated that the alternate codon for $N^{16}$ (AAC) is present in only a small minority of FMDV genomes and CCC is also a minor population of the codons used for residue $P^{17}$. The results presented in Table 1

clearly indicate that infectious FMDVs with these synonymous changes can be obtained but these viruses do not appear to be stably maintained in cell culture and are apparently selected against.

The evidence presented here strongly suggests that there is a distinct selection, within the virus when grown in cell culture, for codon AAC for $N^{16}$, CCT for $P^{17}$, GGG for $G^{18}$ and CCC for $P^{19}$; thereby indicating that synonymous codon usage for this conserved motif is biased in these rescued viruses. It is particularly noteworthy that the codon preference for $P^{17}$ and $P^{19}$ is different (CCT and CCC respectively). This raises the question of why does the virus select some codons over others? Various studies have demonstrated that synonymous codon usage bias plays an important role in the translation of certain mRNAs (Akashi 2001; Bulmer 1991; Novoa and Ribas de Pouplana 2012; Mauro and Chappell 2014). It is therefore conceivable that synonymous codons may influence the cleavage efficiency through the FMDV StopGo mechanism. As indicated above, a marked codon bias within the FMDV genome is apparent from the alignment of diverse FMDV 2A sequences as described by Gao et al. (2014). However, in the context of a synthetic reporter polyprotein, assayed within CHO cells, use of the four different synonymous codons for residue $G^{18}$ of the 2A peptide resulted in very similar apparent "cleavage" efficiencies at the 2A/2B junction. This was interpreted as showing that it is the amino acid residue rather than the nt sequence which is critical for achieving cleavage (Gao et al. 2014). However, using that assay system, the "cleavage" efficiency was only about 88-89% while essentially 100% cleavage occurs within the native context, as in the virus. The results obtained here (see Table 1) indicate that two separate selection effects may be operating. There is a clear selection for the NPGP motif at the amino acid level. However, in addition, there is a distinct codon bias within the context of the rescued infectious viruses and a significant selection pressure appears to exist for the wt sequence. This effect is fully consistent with the codon bias observed in the analysis of

natural FMDV genomic sequences (Gao et al. 2014). This suggests that the FMDV RNA sequence itself (rather than just the encoded amino acid sequence) affects the "cleavage" process (StopGo mechanism) at the 2A/2B junction. Such an effect could be achieved through a direct interaction of the RNA sequence itself or potentially through interactions with the specific charged tRNAs involved in the translation process. In the case of the $P^{17}$ and $P^{19}$ codons, it is interesting to note that the same type of prolyl tRNA (with an IGG anticodon) has been reported to be used for decoding of the CCC and CCU codons in human cells (no gene for a tRNA that is cognate for CCC was identified, see Mauro and Chappell, 2014). However, in the current database of tRNA sequences from the Lowe laboratory, it appears that in humans, 1 of 23 genes for prolyl tRNAs has a GGG anticodon with 10 copies having an AGG anticodon. In the mouse genome, 1 of 20 genes for the prolyl tRNAs has the GGG anticodon and 8 genes have the AGG anticodon (see the gtrnadb.ucsc.edu database described in Chan and Lowe (2009)). Interestingly, in cattle and pigs (major hosts for FMDV) and also in the rat, there is no gene for a prolyl tRNA with a GGG anticodon. Thus, it is not clear whether a single, post-transcriptionally modified prolyl tRNA recognizes these two Pro codons (at least some of the time) or if different tRNAs are involved in the hamster cells used here. If a single tRNA is involved in recognizing both codons (as in cattle, pigs and rats), then it seems that the RNA sequence itself must be influencing the StopGo process; it seems unlikely that this effect is mediated through some secondary or tertiary RNA structure, as this would presumably be lost on the ribosome during the process of translation. It will clearly be important to analyse the effect of the presence of the non-optimal synonymous codons on "cleavage" at the 2A/2B junction in its native context.

## Materials and Methods

### Construction of plasmids containing full-length mutant FMDV cDNAs

Pools of StopGo cDNA mutants that potentially result in all possible single amino substitutions in place of the $N^{16}$, $P^{17}$, $G^{18}$ and $P^{19}$ residues, respectively, were constructed. This was achieved using a 2-step site-directed mutagenesis procedure. This is a variation of the QuickChange protocol (Stratagene), using Phusion High-Fidelity DNA Polymerase (Thermo Scientific) with modified versions of the plasmid pT7S3 (Ellard et al. 1999) as template. The wt pT7S3 contains the full-length cDNA for the O1Kaufbeuren B64 strain of FMDV. To eliminate the possibility of carrying over some residual wt template from the PCR, the templates used were modified versions of the pT7S3 with the codons for $N^{16}$, $P^{17}$, $G^{18}$ or $P^{19}$ changed to encode an alanine (A) residue in each case (see Figure 1A). These substitutions have been reported previously to result in a complete loss of apparent cleavage activity (Sharma et al. 2012; Donnelly et al. 2001b) and it has not been possible to rescue infectious virus containing these substitutions (Kjær J and Belsham GJ, submitted). The first round of PCRs used the forward mutagenic 2A PCR primers (Table 3), with a single reverse primer 8APN206 (Table 3) plus the four different modified pT7S3 plasmids as templates and generated amplicons of ca. 450 bp. These primary PCR products were then used as megaprimers for a second round of PCR with the respective mutant pT7S3 plasmids as templates to produce full-length plasmids. Following *DpnI* digestion, the products from each reaction were introduced into *E. coli* and grown as separate pools. The plasmid pools were sequenced using a BigDye Terminator v. 3.1 Cycle Sequencing Kit and a 3500 Genetic Analyzer (Applied Biosystems).

**Rescue of virus from full-length cDNA plasmids**

Plasmid DNA isolated from each pool was linearized by digestion with *HpaI* and RNA transcripts were prepared using T7 RNA polymerase (Ambion T7 MEGAscript) at 37°C for 4 hours. The integrity of the transcripts was assessed on agarose gels and quantified by

spectrophotometry (NanoDrop 1000, Thermo Scientific) after which they were introduced into BHK cells by electroporation, as described previously (Nayak et al. 2005). The BHK cells were maintained in Dulbecco's Modified Eagles Medium (DMEM) supplemented with 5% fetal calf serum, and incubated at 37°C with 5% $CO_2$. At 2 days post-electroporation, the viruses were harvested by freezing and then amplified through three passages (p2, p3 and p4) in BHK cells.

**Characterization of viruses following multiple passages**

After each passage, viral RNA was extracted from a sample of the virus harvest (using the RNeasy Mini Kit, Qiagen) and converted to cDNA using ready-to-go you-prime first-strand beads (GE Healthcare Life Sciences). FMDV cDNA, which included the whole 2A coding region, was amplified in PCRs (AmpliTaq Gold DNA polymerase, Thermo Scientific) using primers 8APN206 and 8APN203 (see Figure 1 and Table 3). Control reactions, without RT, were used to ensure that the analysed products were derived from RNA and not from the presence of carryover plasmid DNA template. The amplicons (ca. 650 bp) were visualized on 1% agarose gels and purified (GeneJET gel extraction kit, Thermo Scientific). These amplicons should be representative of the heterogeneity present in the rescued virus populations. The resulting collections of fragments were inserted into pCR-XL-TOPO (Thermo Scientific) and the sequence of the cDNA fragment present in individual bacterial clones (20 colonies for each of the 4 residues) was determined using the same reverse primer as used for the PCR. The fragments from codon mutants $G^{18}$ and $P^{19}$ were also enriched for the non-wt sequence populations by digestion of the cDNA with *ApaI* prior to gel purification and insertion into the pCR-XL-TOPO vector as described above.

## Acknowledgements

We thank Preben Normann for excellent technical assistance. The studies were supported by core funds within DTU-Vet.

## References


Akashi H. 2001. Gene expression and molecular evolution. *Curr Opin Genet Dev* **11**: 660–666.

Atkins JF, Wills NM, Loughran G, Wu C-Y, Parsawar K, Ryan MD, Wang C-H, Nelson CC. 2007. A case for "StopGo": reprogramming translation to augment codon meaning of GGN by promoting unconventional termination (Stop) after addition of glycine and then allowing continued translation (Go). *RNA* **13**: 803–810.

Belsham GJ. 2005. Translation and replication of FMDV RNA. *Curr Top Microbiol Immunol* **288**: 43–70.

Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**: 897–907.

Chan PP, Lowe TM. 2009. GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucl Acids Res* **37**:D93-D97.

Donnelly ML, Gani D, Flint M, Monaghan S, Ryan MD. 1997. The cleavage activities of aphthovirus and cardiovirus 2A proteins. *J Gen Virol* **78**: 13–21.

Donnelly ML, Luke G, Mehrotra A, Li X, Hughes LE, Gani D, Ryan MD. 2001a. Analysis of the aphthovirus 2A/2B polyprotein "cleavage" mechanism indicates not a proteolytic reaction, but a novel translational effect: A putative ribosomal "skip." *J Gen Virol* **82**: 1013–1025.

Donnelly ML, Luke GA, Hughes LE, Luke G, Mendoza H, Dam E, Gani D, Ryan MD. 2001b. The "cleavage" activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring "2A-like" sequences. *J Gen Virol* **82**: 1027–1041.

Doronina VA, Wu C, de Felipe P, Sachs MS, Ryan MD, Brown JD. 2008. Site-specific release of nascent chains from ribosomes at a sense codon. *Mol Cell Biol* **28**: 4227–4239.

Ellard FM, Drew J, Blakemore WE, Stuart DI, King AMQ. 1999. Evidence for the role of His 142 of protein 1C in the acid induced disassembly of foot and mouth disease virus capsids. *J Gen Virol* **80**: 1911–1918.

Gao ZL, Zhou JH, Zhang J, Ding YZ, Liu YS. 2014. The silent point mutations at the

cleavage site of 2A/2B have no effect on the self-cleavage activity of 2A of foot-and-mouth disease virus. *Infect Genet Evol* **28**: 101–106.

Gullberg M, Polacek C, Bøtner A, Belsham GJ. 2013. Processing of the VP1/2A junction is not necessary for production of foot-and-mouth disease virus empty capsids and infectious viruses: characterization of "self-tagged" particles. *J Virol* **87**: 11591–11603.

Hahn H, Palmenberg AC. 1996. Mutational analysis of the encephalomyocarditis virus primary cleavage. *J Virol* **70**: 6870–6875.

Kristensen T, Normann P, Gullberg M, Fahnøe U, Polacek C, Rasmussen TB, Belsham GJ. 2017. Determinants of the VP1/2A junction cleavage by the 3C protease in foot-and-mouth disease virus-infected cells. *J Gen Virol* **98**: 385–395.

Loughran G, Libbey JE, Uddowla S, Scallan MF, Ryan MD, Fujinami RS, Rieder E, Atkins JF. 2013. Theiler's murine encephalomyelitis virus contrasts with encephalomyocarditis and foot-and-mouth disease viruses in its functional utilization of the StopGo non-standard translation mechanism. *J Gen Virol* **94**: 348–353.

Martinez-Salas E, Belsham GJ. 2017. Genome Organisation, Translation and Replication of Foot-and-mouth Disease Virus RNA. In *Foot-and-Mouth Disease: Current Research and Emerging Trends* (eds. F. Sobrino and E. Domingo), pp. 13–42, Caister Academic Press.

Mauro VP, Chappell SA. 2014. A critical analysis of codon optimization in human therapeutics. *Trends Mol Med* **20**: 604–613.

Nayak A, Goodfellow IG, Belsham GJ. 2005. Factors required for the Uridylylation of the foot-and-mouth disease virus 3B1, 3B2, and 3B3 peptides by the RNA-dependent RNA polymerase (3Dpol) in vitro. *J Virol* **79**: 7698–7706.

Novoa EM, Ribas de Pouplana L. 2012. Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends Genet* **28**: 574–581.

Palmenberg AC, Parks GD, Hall D, Ingraham RH, Seng TW, Pallal P V. 1992. Proteolytic processing of the cardioviral cleavage in clone-derived P2 region□: primary 2A / 2B precursors. *Virology* **190**: 754–762.

Ryan MD, Donnelly M, Lewis A, Mehrotra AP, Wilkie J, Gani D. 1999. A Model for Nonstoichiometric, Cotranslational Protein Scission in Eukaryotic Ribosomes. *Bioorg*

*Chem* **27**: 55–79.

Ryan MD, Drew J. 1994. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J* **13**: 928–33.

Ryan MD, King AMQ, Thomas GP. 1991. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. *J Gen Virol* **72**: 2727–2732.

Sharma P, Yan F, Doronina VA, Escuin-Ordinas H, Ryan MD, Brown JD. 2012. 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucleic Acids Res* **40**: 3143–3151.

Tulloch F, Luke GA, Ryan MD. 2017. Foot-and-mouth disease virus proteinases and polyprotein processing. In *Foot-and-Mouth Disease: Current Research and Emerging Trends* (eds. F. Sobrino and E. Domingo), pp. 43–59, Caister Academic Press.

**Table 1: Codon utilization encoding the "NPGP" motif at the 2A/2B junction within rescued FMDVs.**

| Residue | Codon | | p2 %[1] | p3 %[1] | p4 %[1] |
|---|---|---|---|---|---|
| N16 | *AAT* | | 45 | 25 | 5 |
| N16 | AAC | wt | 55 | 75 | 95 |
| P17 | CCT | wt | 55 | 70 | 100 |
| P17 | *CCC* | | 10 | 15 | 0 |
| P17 | *CCA* | | 20 | 5 | 0 |
| P17 | *CCG* | | 15 | 10 | 0 |
| G18 | *GGT* | | 15 | 5 | 0 |
| G18 | *GGC* | | 15 | 5 | 0 |
| G18 | *GGA* | | 50 | 40 | 10 |
| G18 | GGG | wt | 20 | 50 | 90 |
| P19 | *CCT* | | 50 | 25 | 5 |
| P19 | CCC | wt | 10 | 50 | 95 |
| P19 | *CCA* | | 10 | 5 | 0 |
| P19 | *CCG* | | 30 | 20 | 0 |

**1**: From sequencing of plasmid DNA isolated from separate 20 colonies in each case, the proportion (%) of each codon present in the rescued FMDVs is indicated at the different passage (p) numbers. Codon frequency values of 50-70% are highlighted in light grey whereas values from 75-100% are highlighted in dark grey.

345 **Table 2: Enrichment for non-wt sequences encoding residues G[18] and P[19] within the**
346 **"NPGP" motif within rescued FMDVs.**

| Residue | Codon | | Pretreatment | p2 %[2] | p3 %[2] |
|---|---|---|---|---|---|
| G18 | *GGT* | | *ApaI*[1] | 5 | 0 |
| G18 | *GGC* | | *ApaI*[1] | 15 | 5 |
| G18 | *GGA* | | *ApaI*[1] | 80 | 95 |
| G18 | GGG | wt | *ApaI*[1] | 0 | 0 |
| P19 | *CCT* | | *ApaI*[1] | 60 | 55 |
| P19 | CCC | wt | *ApaI*[1] | 0 | 0 |
| P19 | *CCA* | | *ApaI*[1] | 15 | 0 |
| P19 | *CCG* | | *ApaI*[1] | 25 | 45 |

347

348 **1**: Following RT-PCR, the 650bp amplicons were digested with *ApaI* to enrich the population
349 in non-wt sequences and the residual intact products were inserted into the pCR-XL-TOPO
350 vector (see text).

351 **2**: From sequencing of plasmid DNA isolated from separate 20 colonies in each case, the
352 proportion (%) of each codon present in the rescued FMDVs is indicated at the different
353 passage (p) numbers. Codon frequency values of 50-70% are highlighted in light grey
354 whereas values from 75-100% are highlighted in dark grey.

355

356

357

358 **Table 3: Primers used to create and sequence mutant FMDV cDNAs.**

| Primer | Sequence (5'-3') |
|---|---|
| Fwd_2A_N16A_degen | GGAGTCCNNNCCTGGGCCCTTC |
| Fwd_2A_P17A_ degen | GTCCAACNNNGGGCCCTTC |
| Fwd_2A_G18A_ degen | GACGTCGAGTCCAACCCTNNNCCCTTCTTTTTCTCCGACGTTA |
| Fwd_2A_P19G_ degen | TCG AGTCCAACCCTGGGNNNTTCTTTTTCTCCGACGTTAGG |
| 8APN206 | CACCCGAAGACCTTGAGAG |
| 8APN203 | CTCCTTCAACTACGGTGCC |

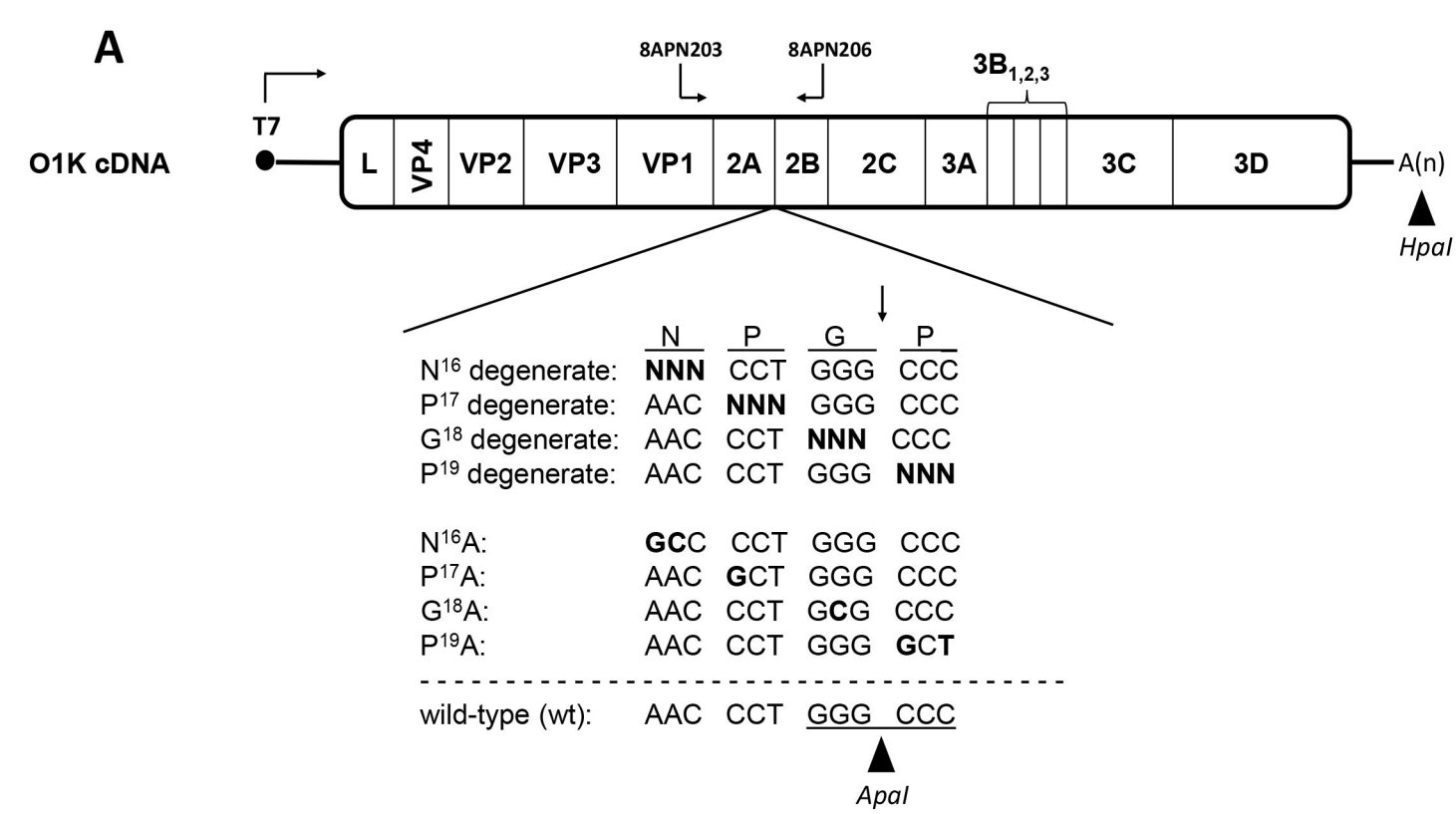
359

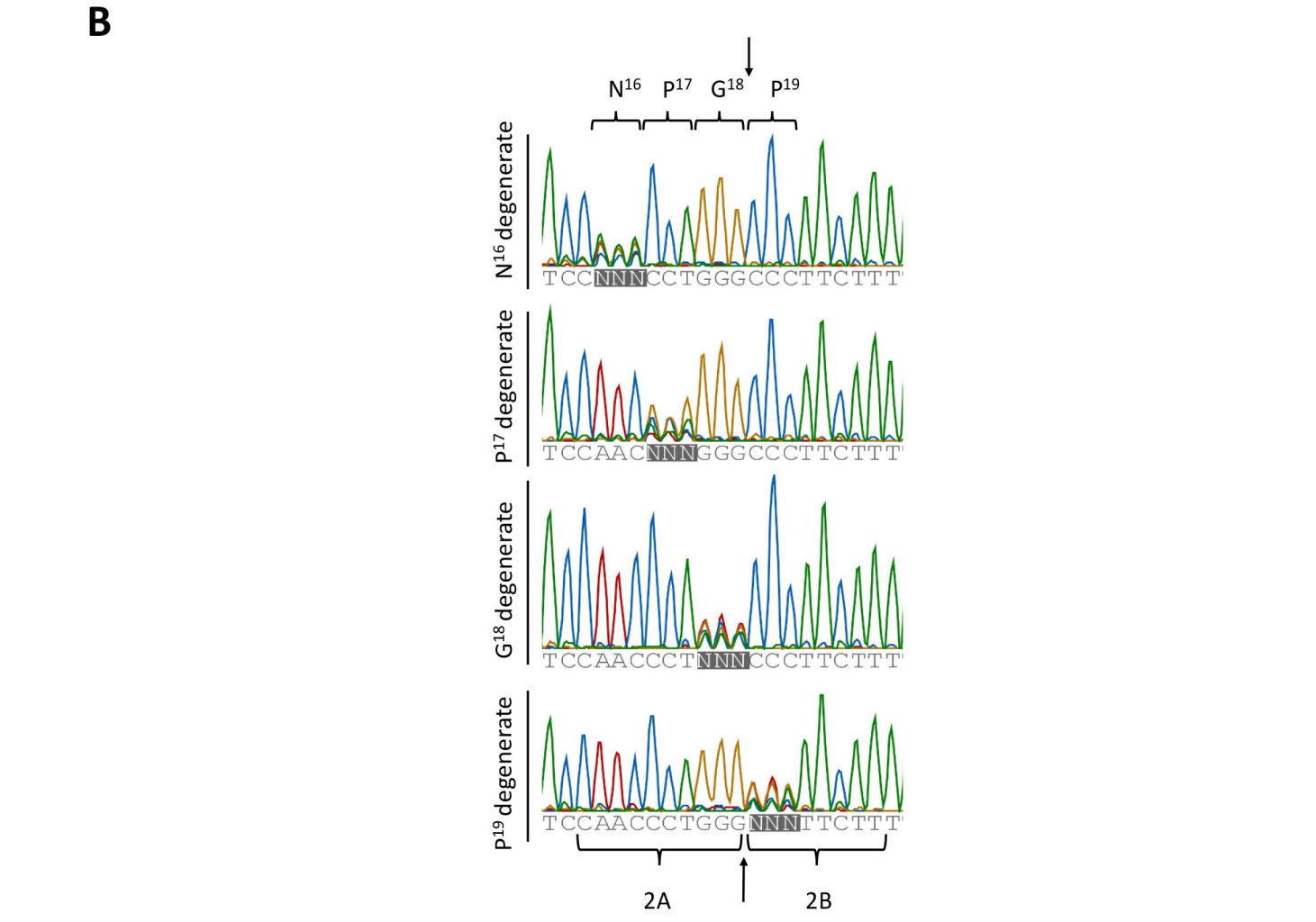
360

**Figure legend**.

**Figure 1: Structure of the FMDV O1 Kaufbeuren (O1K) cDNA and its derivatives.** (A) The plasmid-encoded amino acids and the corresponding nucleotide sequences at the 2A/2B junction are shown. The FMDV O1K degenerate codon mutants were produced as described in the text using the mutant pT7S3 plasmids encoding the $N^{16}A$, $P^{17}A$, $G^{18}A$ and $P^{19}A$ substitutions as templates. The full-length plasmid pools were linearized using *HpaI* prior to *in vitro* transcription and virus rescue. The locations of the *HpaI* and *ApaI* restriction sites that were used are marked. N = a mixture of the 4 nucleotides. (B) Chromatograms and sequences of the FMDV cDNA corresponding to the NPGP motif at the 2A/2B junction. Degenerate positions showing the presence of multiple nucleotides are marked with an N (in bold type). The colour code in the chromatograms is as follows: A (red), T (green), G (yellow), C (blue).

389


A
8APN203
8APN206
3B1,2,3
T7
O1K cDNA
L
VP4
VP2
VP3
VP1
2A
2B
2C
3A
3C
3D
A(n)
HpaI
N P G P
N16 degenerate: NNN CCT GGG CCC
P17 degenerate: AAC NNN GGG CCC
G18 degenerate: AAC CCT NNN CCC
P19 degenerate: AAC CCT GGG NNN
N16A: GCC CCT GGG CCC
P17A: AAC GCT GGG CCC
G18A: AAC CCT GCG CCC
P19A: AAC CCT GGG GCT
wild-type (wt): AAC CCT GGG CCC
ApaI


B
N16 P17 G18 P19
N16 degenerate
TCCNNNCCTGGGCCCTTCTTT
P17 degenerate
TCCAACNNNGGGCCCTTCTTT
G18 degenerate
TCCAACCCTNNNCCCTTCTTT
P19 degenerate
TCCAACCCTGGGNNNTTCTTT
2A
2B


# Selection of functional 2A sequences within foot-and-mouth disease virus; requirements for the NPGP motif with a distinct codon bias

Jonas Kjær and Graham J. Belsham

*RNA* published online October 17, 2017